CODING CAPACITY FOR A CLASS OF ADDITIVE CHANNELS

Charles R. Baker

Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599

LISS-36
May 1989

Abstract

Coding capacity is determined for a class of additive Gaussian channels, and bounds on capacity are obtained for a class of nonGaussian channels. The channels may be with or without memory, stationary or nonstationary. The constraint is partially given in terms of an increasing family of finite-dimensional subspaces. A general expression for the capacity is obtained, which depends upon the relation between the noise covariance and the constraint on the generalized signal-to-noise energy ratio for the code words. The well-known expression for capacity of the discrete-time stationary Gaussian channel is shown to be a special case. The general expression provides new results on capacity for nonstationary discrete-time channels and for continuous-time channels (stationary or nonstationary) with fixed time of transmission.

This research was supported by ONR Contract N00014-86-K-0039, ONR Grant N00014-89-J-1175 and NSF Grant NCR8713726.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0039 and N00014-89-J-1175
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
NR	042	269	SRO 105

11. TITLE (Include Security Classification)
Coding Capacity for a Class of Additive Channels

12. PERSONAL AUTHOR(S) C.R. Baker

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM ______ TO ______	May 1989	35

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Coding capacity, Additive Gaussian channels, Nonstationary discrete-time channels, Continuous-time channels

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Coding capacity is determined for a class of additive Gaussian channels, and bounds on capacity are obtained for a class of nonGaussian channels. The channels may be with or without memory, stationary or nonstationary. The constraint is partially given in terms of an increasing family of finite-dimensional subspaces. A general expression for the capacity is obtained, which depends upon the relation between the noise covariance and the constraint on the generalized signal-to-noise energy ratio for the code words. The well-known expression for capacity of the discrete-time stationary Gaussian channel is shown to be a special case. The general expression provides new results on capacity for nonstationary discrete-time channels and for continuous-time channels (stationary or nonstationary) with fixed time of transmission.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED XX SAME AS RPT. XX DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
C.R. Baker	(919) 962-2189	

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE. UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Introduction

Coding capacity of additive Gaussian channels with memory is one of the major areas of open problems in basic information theory. Even for the case of the stationary discrete-time channel with a simple energy constraint, only recently has a complete proof been given [15] for the information capacity, which one can then apply toward a rigorous and complete proof of the coding capacity. For nonstationary discrete-time and continuous-time channels with or without memory, there are apparently no published results on coding capacity.

Moreover, in the classical continuous-time channel, the model for which results have been known constitutes a proper subset of the class of stationary channels, and there is a very large universe of stationary channels not belonging to this subset.

This paper gives results on coding capacity for a large class of channels, which may be stationary or nonstationary, with or without memory. The formulation is somewhat different and more general than that usually followed. The generality permits one to focus on channels where dimensionality of the code word set is a key component of the constraints. In the classical setup, the elements of a code are limited in their time duration. The present paper replaces this with a constraint defined by an increasing family of finite-dimensional subspaces. The classical discrete-time channel is then a special case of this framework, and several applications to these channels are given. These applications include nonstationary single-user and multi-user channels. For example, it will be seen that this formulation shows that it is possible to use a code word set of arbitrarily large cardinality as transmission time $n \to \infty$, with the maximum decoding error probability converging to zero, while the classical analysis gives zero capacity and a maximum decoding error of one for any non-zero rate. Another interesting result, for the memoryless non-stationary Gaussian channel, is that the noise covariance can have eigenvalues of infinite multiplicity which have no effect on coding capacity.

The approach also provides results for continuous-time channels. In the classical continuous-time channel, the transmission time T is permitted to be arbitrarily long in determining capacity. Then, by transmitting at a rate below capacity and for a sufficiently long time, the coder has the ability to use an arbitrarily large code word set while achieving arbitrarily small maximum decoding error probability. In this formulation, transmission rate is the rate of increase in the log of the cardinality of the code word set, as the transmission time is increased.

Suppose, however, that the transmission time T is limited, as will ordinarily be the case in practice. One may then ask: if arbitrarily large transmission energy is available, is it possible to choose a code word set of arbitrarily large cardinality while achieving arbitrarily small maximum decoding error probability? The mechanism for accomplishing this, if it is possible, will consist of using an increasingly-complex coding-decoding structure. This can be interpreted as an increase in dimensionality of the code word set. The "rate" of transmission is now the rate of increase one obtains in the log of the cardinality of the code word set as its dimensionality is increased. A higher rate implies that the coding-decoding structure can be less complex for a specified cardinality of the code word set and a specified maximum error probability.

It is shown here that for some such channels it is not possible to achieve arbitrarily small maximum decoding error probability when using code word sets of arbitrarily large cardinality, for any positive rate. A special case of these channels is the Holsinger-Gallager model of the stationary Gaussian channel analyzed in [11] (when the time duration is fixed). Moreover, for such channels it is shown that any non-zero rate leads to a maximum decoding error probability of one. However, for a large class of continuous-time channels of fixed time duration, it is possible to achieve arbitrarily small decoding error probability with code word sets of arbitrarily large cardinality; those channels are characterized, a number of examples are given, and their capacity is obtained.

Upper bounds on coding capacity are obtained for a large class of nonGaussian channels. Several examples are included. For the class of channels considered, it is shown that coding capacity is equal to information capacity when the noise is Gaussian. Apparently, this has only recently been explicitly stated for the classical discrete-time channel (with memory) [8].

Emphasis here is on obtaining the capacity. However, Theorem 2 gives bounds on error probability for Gaussian channels based on results of Ebert [10] and Gallager [11].

In addition to obtaining specific new results for coding capacity of a large class of additive channels, the development brings out the essential importance to the capacity of the limit points of the spectrum (the essential spectrum) of the operator defining the relationship between the noise covariance and the energy constraint on the code words.

The proof of the general expression for the capacity is based on the spectral theory for self-adjoint operators in Hilbert space, including the integral representation (as given, for example, in [17]). That proof, and those of several other necessary mathematical results, is contained in [6]. The emphasis here is on applications.

Problem Framework

In the next few sections, the setting and definitions for the coding capacity problem will be given. In order to illustrate these concepts and definitions, the classical discrete-time channel will be frequently employed.

It is assumed that the noise sample paths belong to a real separable Hilbert space H, where H has inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$. The noise is described by a set function μ_N. μ_N will be a finitely-additive probability defined on the cylinder sets of H: the collection of all sets of the form $\{x: (\langle x,u_1\rangle,\ldots,\langle x,u_n\rangle) \in D_n\}$, where $n \geq 1$, D_n is a Borel set in $\mathbb{R}^n$, and $u_1,\ldots,u_n$ are any n elements of H. Thus, if H_0 is any finite-dimensional subspace of H, and P_0 is the projection operator in H having range H_0, let μ_0 be defined on the Borel sets of H_0 by $\mu_0(A) = \mu_N\{x: P_0x \in A\}$. μ_0 is then a countably-additive probability. Our basic assumptions on the noise are that

a) $\int_H \langle x,y\rangle^2 d\mu_N(x) < \infty$ all y in H, and

b) $\int_H \langle x,y\rangle d\mu_N(x) = 0$ for all y in H.

Assumption (a) means that μ_N has a covariance operator R_N, which is linear, bounded, and non-negative, and also implies that the noise mean exists. Assumption (b) is that the noise has zero mean. We can assume WLOG (without loss of generality) that R_N is strictly positive on H. R_N is defined as $\langle R_N u,v\rangle = \int_H \langle u,x\rangle\langle v,x\rangle d\mu_N(x)$ for u,v in H.

μ_N is Gaussian if for any $y \in H$ the distribution function P_y is Gaussian, $P_y(\alpha) = \mu_N\{x: \langle x,y\rangle \leq \alpha\}$.

Formulation of Constraints

Let R_W be a strictly positive covariance operator in H satisfying $\text{range}(R_W^{1/2}) \subset \text{range}(R_N^{1/2})$.

(H_n), $n \geq 1$, will denote a family of finite-dimensional subspaces of H, such that for all $n \geq 1$,

a) $H_n \subset H_{n+1}$

b) $\dim(H_n) = n$.

Let $\tilde{P}_n$ be the projection operator with range equal to H_n, and define $R_{W,n} = \tilde{P}_n R_W \tilde{P}_n$. $R_{W,n}$ is then strictly positive on H_n. For x in H_n, the norm $\|x\|_{W,n}$ is defined by $\|x\|_{W,n} = \|y\|$, where y is the unique element of H_n that satisfies $R_{W,n}^{1/2} y = x$. Formally, we write $\|x\|_{W,n} = \|R_{W,n}^{-1/2} x\|$ for x in H_n; although $R_{W,n}^{-1/2}$ does not exist on H, it is well-defined on H_n.

The constraints on the code words are now as follows: For each $n \geq 1$, the admissible code words $x_1, \ldots, x_{K(n)}$ belong to H_n and satisfy $\|x_i\|_{W,n}^2 \leq nP$ for $i = 1,2,\ldots,K(n)$.

As an important example of such constraints, consider the classical discrete-time memoryless channel with a simple energy constraint. This can be formulated in the above terms by taking $H = \ell_2$ and R_W the identity, giving

$$H_n = \{x \text{ in } \ell_2\colon\ x_i = 0 \text{ for } i > n\}$$

$$\|x\|_{W,n}^2 = \sum_{i=1}^{n} x_i^2 \qquad \text{for } x \text{ in } H_n.$$

Another example involves the continuous-time channel with fixed transmission time, T. In this case, H can be taken as $L_2[0,T]$, and H_n as the linear span of $\{e_1, e_2, \ldots, e_n\}$, where $\{e_n, n \geq 1\}$ is an infinite orthonormal set in $L_2[0,T]$. It will always be assumed that a process with paths in $L_2[0,T]$ is product-measurable.

It can be noted that for channel capacity calculations the constraint $\|x\|_{W,n}^2 \leq nP$ (for code words in H_n) for every n is equivalent to the constraint $\limsup_{n\to\infty} \frac{1}{n} \|x\|_{W,n}^2 \leq P$.

Definition of Coding Capacity

Let $(H_n, \|\cdot\|_{W,n})$ define the constraints. Let μ_N be the noise probability on H, and let μ_N^n be the probability on H_n induced from μ_N by the projection operator $\tilde{P}_n$: $\mu_N^n[A] = \mu_N\{x: \tilde{P}_n x \in A\}$, for A any Borel set in H_n. It is not required that μ_N be countably additive; thus, for example, the discrete-time memoryless stationary Gaussian channel is included in this formulation.

For fixed $n \geq 1$, a *code* (k,n,ϵ) [1] is a set of k code words $\{x_1,\ldots,x_k\}$ and k disjoint Borel sets in H_n such that the elements of $\{x_1,\ldots,x_k\}$ obey the constraints, and

$$\mu_N\{y: \tilde{P}_n y + x_i \in C_i\} \geq 1 - \epsilon \text{ for } i = 1,2,\ldots,k.$$

Note that this last probability inequality can be written as

$$\mu_N^n\{y: y + x_i \in C_i\} \geq 1 - \epsilon, \quad i = 1,2,\ldots,k.$$

A real number $R \geq 0$ is an admissible rate if there exists an infinite sequence of codes $([e^{n_i R}], n_i, \epsilon_{n_i})$ with $\epsilon_{n_i} \to 0$ as $i \to \infty$, where $[r]$ is the integer part of the real number r.

The coding capacity is then the supremum of the set of admissible rates.

As can be seen, this definition of coding capacity contains that of the classical discrete-time channel as a special case, defining the constraint *family* $(H_n, \|\cdot\|_{W,n})$ as in the previous section. More generally, the capacity gives an indication of the effect on size of the code word set that can be obtained by increasing the dimensionality of the code word set, while requiring that $\liminf_{n\to\infty} \epsilon_n = 0$.

The constraint on the transmitted signal is given in terms of a covariance operator R_W in H. A basic assumption is that range$(R_N^{\frac{1}{2}})$ contains range$(R_W^{\frac{1}{2}})$. The existence of such an operator, and the assumption on range relations, are necessary in order that the information capacity be finite [5]; moreover, when μ_N is Gaussian, it will be seen that finite information capacity is necessary in order to have finite coding capacity. Thus, the formulation of the problem is quite general (so long as an average-type constraint is to be used). Under this assumption, there exists a self-adjoint operator S in H such that $R_N = R_W^{\frac{1}{2}}(I+S)R_W^{\frac{1}{2}}$, where $(I+S)^{-1}$ exists and is bounded (see [5, Prop. 1] for ramifications of this fact). The limit points of the spectrum of S will play a key role in this paper. These limit points (the essential spectrum of S) consist of all eigenvalues of infinite multiplicity, all limits of sequences of distinct eigenvalues, and all points of the continuous spectrum [17, p. 363]. "Essential spectrum" is the modern terminology for this set; it will be denoted by $\sigma_{ess}(S)$. The continuous spectrum of S consists of all real numbers λ such that the range of $(S-\lambda I)$ is not closed.

In many applications, the constraint will be given by a time-invariant linear filter f with transfer (frequency) function $\hat{f}$. In such cases, $|\hat{f}|^2$ defines a spectral density and thereby the operator R_W.

<u>Expressions for Evaluation of the Coding Capacity</u>

The noise has probability μ_N and covariance operator R_N. μ_N^n is the noise probability on H_n, defined by $\mu_N^n(C) = \mu_N\{x: \tilde{P}_n x \in C\}$ for C a Borel set in H_n. R_W is the covariance operator defining the constraint on the code words. $R_{W,n} = \tilde{P}_n R_W \tilde{P}_n$ and $R_{N,n} = \tilde{P}_n R_N \tilde{P}_n$. Let I_n be the identity in H_n; let S_n be the self-adjoint operator mapping H_n into H_n defined by $R_{N,n} = R_{W,n}^{\frac{1}{2}}(I_n+S_n)R_{W,n}^{\frac{1}{2}}$. $\|S_n x\| \equiv 0$ if x is orthogonal to H_n. Since S_n is self-adjoint as an operator in H_n, it has n orthonormal eigenvectors belonging to H_n with corresponding eigenvalues $\beta_1^n \le \beta_2^n \le \dots \le \beta_n^n$.

Absolute continuity of probabilities will be frequently encountered. If

μ and υ are two finitely-additive functions on the cylinder sets of H, then $\mu << \upsilon$ and $\mu \sim \upsilon$ will denote, respectively, absolute continuity of μ w.r.t. υ and mutual absolute continuity of μ and υ. $\mu << \upsilon$ if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\upsilon(A) < \delta \Rightarrow \mu(A) < \epsilon$ for any cylinder set A in H.

The bound on coding capacity for nonGaussian channels will involve the relative entropies $H_{GN}(N)$ and $\{H^n_{GN}(N),\ n \geq 1\}$. Let μ_{GN} be the Gaussian noise measure (perhaps not countably additive) having covariance operator R_N. In this framework, the definition is $H_{GN}(N) = \sup_n H^n_{GN}(N)$, where $H^n_{GN}(N)$ is the entropy of μ^n_N with respect to μ^n_{GN}: $H^n_{GN}(N) = \int_{\mathbb{R}^n}\left[\log \frac{d\mu^n_N}{d\mu^n_{GN}}\right]d\mu^n_N$. Of course, $H^n_{GN}(N) = \infty$ if μ^n_N is not absolutely continuous with respect to μ^n_{GN}. $\mu'_N << \mu'_{GN}$ is necessary in order to have $H_{GN}(N)$ finite, where μ' denotes the restriction of μ to the closed linear span of $\bigcup_{n\geq 1} H_n$.

The relative entropy $H^n_{GN}(N)$ can be defined in terms of differential entropy for the discrete-time channel. Suppose that $\underline{N} = (N_1, N_2, \ldots, N_n)$ has zero mean and a probability density p^n_N. Then, $d\mu^n_N/d\mu^n_{GN}$ exists, since Lebesgue measure and nondegenerate Gaussian measure are mutually absolutely continuous on $\mathbb{R}^n$. The differential entropy is

$$H^n(N) = -\int_{\mathbb{R}^n}[\log p^n_N(\underline{x})]p^n_N(\underline{x})d\lambda^n(\underline{x})$$

where λ^n denotes Lebesgue measure on $\mathbb{R}^n$. Thus,

$$H^n(N) = -\int_{\mathbb{R}^n}\left[\log \frac{d\mu^n_N}{d\mu^n_{GN}}(\underline{x})\right]d\mu^n_N(\underline{x}) - \int_{\mathbb{R}^n}\left[\log \frac{d\mu^n_{GN}}{d\lambda^n}(\underline{x})\right]d\mu^n_N(\underline{x})$$

$$= -H^n_{GN}(N) + H^n(GN).$$

so that $H^n_{GN}(N) = H^n(GN) - H^n(N)$.

The rel. ive entropy will enable us to give an upper bound for nonGaus- an channels such that $\overline{\lim_n} \frac{1}{n} H_{GN}(N) < \infty$. The bound will be seen to be equal to the capacity for the Gaussian channel with noise probability μ_{GN} whenever $\overline{\lim_n} \frac{1}{n} H^n_{GN}(N) = 0$. A particular case of this is when $H_{GN}(N) < \infty$. To illustrate that this occurs in some important applications, suppose that $H = L_2[0,T]$ and that μ_N is defined by a mean-square continuous stochastic process (N_t). Suppose also that $(N_t) = (V_t+S_t)$, where (V_t) is a m.s. continuous Gaussian process and (S_t) is a process independent of (V_t) and such that the paths of (S_t) belong (w.p. 1) to the RKHS of the covariance of (V_t). Suppose also that the Gaussian process with the (S_t) covariance has sample paths in the RKHS of the covariance of (V_t), w.p. 1. Then, $H_{GN}(N) < \infty$.

This result is a special case of the following.

Prop. 1. For any choice of (H_n), $H_{GN}(N) < \infty$ in each of (a) - (d) below.

(a) μ_V is Gaussian with covariance R_V, μ_S has covariance $R_S = R_V^{1/2} T R_V^{1/2}$ for T trace-class, and $\mu_N = \mu_V * \mu_S$ (convolution).

(b) $H = \ell_2$ or $H = L_2[0,T]$, V is a Gaussian process with sample paths a.s. in H and covariance operator R_V, S is a possibly non-Gaussian process independent of V with sample paths a.s. in H and with covariance operator R_S, N = S + V, and $R_S = R_N^{1/2} T R_N^{1/2}$ for T trace-class.

(c) V, S, and N are as in (b), S' is the Gaussian process with the same covariance function as S, and the paths of S' belong to range$(R_V^{1/2})$ with probability 1.

(d) $H = L_2[0,T]$, S, V, and N are defined as in (b), V and S are wide-sense stationary and have rational spectral densities Φ_V and Φ_S, and

$$\int_{-\infty}^{\infty} \frac{\Phi_S}{\Phi_V}(\lambda)d\lambda < \infty.$$

Proof. (b) and (c) are equivalent [2]. (d) is a special case of (b) [2, 13]. (b) is obviously a special case of (a). The proof will be given for (a), writing μ_N as μ_{S+V}. Let $\mu_V \circ f_x^{-1}(A) = \mu_V\{y: x+y \in A\}$, where $x \in H$ and $f_x(y) \equiv x + y$. The following statements follow from [2]:

a) $\mu_V \circ f_x^{-1} \sim \mu_V$ a.e. $d\mu_S(x)$

b) $\mu_V \sim \mu_N$

c) $\mu_N \sim \mu_{GN}$.

Consider now the channel with additive Gaussian noise μ_V, and let its information capacity C(P) be sup $I(\mu_{XY})$, where the sup is over all probabilities μ_X such that $\mu_X[\text{range}(R_N^{1/2})] = 1$ and $E_{\mu_X} \|x\|_N^2 \leq P$, where $\|x\|_N \equiv \|R_N^{-1/2}x\|$. The map g: $(x,y) \to [(d\mu_V \circ f_x^{-1})/d\mu_V](y)$ is $\overline{\mathbb{B}[H]\times\mathbb{B}[H]}/\mathbb{B}[\mathbb{R}]$ measurable [4]. For any μ_X satisfying the constraint, we have [3] $I(\mu_{XY}) = \frac{1}{2}\,\text{Trace}\, R_V^{-1/2}R_XR_V^{-1/2} - H_V(X+V)$. Trace $R_V^{-1/2}R_XR_V^{-1/2} \leq P$, from the constraint. Since $I(\mu_{XY}) \geq 0$, this requires $H_V(X+V) < \infty$. Finally, since $N = X + V$,

$$H_{GN}(N) = \int\left[\log \frac{d\mu_N}{d\mu_{GN}}\right]d\mu_N = \int\left[\log \frac{d\mu_N}{d\mu_V} \cdot \frac{d\mu_V}{d\mu_{GN}}\right]d\mu_N$$

$$= H_V(N) + \int\left[\log \frac{d\mu_V}{d\mu_{GN}}\right]d\mu_N \leq H_V(N) + \int\left[\log \frac{d\mu_V}{d\mu_{GN}}\right]d\mu_V$$

Since μ_V and μ_{GN} are mutually absolutely continuous and Gaussian, $H_{GN}(V) < \infty$. Hence, $H_{GN}(N) < \infty$. □

The model just described arises in one of the most frequently-encountered nonGaussian situations: when the medium contains additive narrowband nonGaussian noise (S_t) that is independent of the additive wideband Gaussian receiver noise (V_t). The above case applies, for example, if the receiver noise is stationary with spectral density $\Phi(\lambda) = 1/(\lambda^2+\alpha^2)$, and the ambient medium noise is stationary and nonGaussian with spectral density $\Phi_1(\lambda) = 1/B(\lambda)$, where $B(\lambda)$ is a polynomial of order ≥ 4.

Consider now the finite-dimensional channel defined as follows. The additive noise has probability μ_N^n. The input to the channel is described by a probability μ_X on H_n, satisfying $E_{\mu_X} \|x\|^2_{W,n} \leq nP$. Let $C_W^n(nP)$ be the information capacity of this channel. The following well-known result is fundamental to our results.

Lemma 1 [11, 14].

$$\frac{1}{2} \sum_{i=1}^{N(n)} \log\left[\frac{B(n)+1}{\beta_i^n+1}\right] \leq C_W^n(nP) \leq \frac{1}{2} \sum_{i=1}^{N(n)} \log\left[\frac{B(n)+1}{\beta_i^n+1}\right] + H_{GN}^n(N)$$

where $\beta_1^n \leq \beta_2^n \leq \ldots \leq \beta_n^n$ are the eigenvalues of S_n, $N(n) = \sup\{i \leq n: \beta_i^n \leq B(n)\}$, and $B(n)$ is defined by

$$P = \frac{1}{n} \sum_{i=1}^{N(n)} (B(n) - \beta_i^n).$$

Preliminary Results

Our program is to first obtain expressions for $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$. This will then be followed by the result that the coding capacity is equal to $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$ when μ_N is Gaussian and that this value is an upper bound for coding capacity for the nonGaussian processes satisfying $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$. Among the difficulties in evaluation of $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$ is that for each value of n one obtains an expression in terms of the eigenvalues of the operator S_n. The range of S_n is contained in H_n; S_n always has a complete orthonormal set of eigenvectors, and its range space has dimension $\leq n$. Moreover, the eigenvalues of S_n need not be contained in the set of eigenvalues of S_{n+1}; in fact, the eigenvalues of S_{n+1} may not include a single eigenvalue of S_n.

The desired result is an expression for capacity in terms of the operator S, where $R_N = R_W^{½}(I+S)R_W^{½}$, and the increasing subspaces (H_n). This requires that one first determine relations between S and S_n, and in particular, between the spectral properties of the two operators. Some idea of the complexity of this procedure may be gained by observing that, in general, S need not have any eigenvectors. Examples of such S include the following: $H = \ell_2$ (discrete-time channel) with S a Toeplitz matrix; $H = L_2[0,T]$, with $[Sx](t) \equiv t^r x(t)$ a.e. dt, some real number $r > 0$.

In [6], it is shown that $S_n = V_n S V_n^*$, where V_n is a partially isometric operator. V_n is isometric on range$(R_W^{½}\tilde{P}_n)$, zero on $[\text{range}(R_W^{½}\tilde{P}_n)]^{\perp}$, and its range is equal to H_n. Let $H_{W_n} \equiv \text{range}(R_W^{½}\tilde{P}_n)$. The eigenvalues (β_i^n) of S_n and their multiplicity are then the same as those of the operator $P_{W_n} S P_{W_n}$, where P_{W_n} is the projection operator with range equal to H_{W_n} [6].

$\{e_i, i \geq 1\}$ will be used to denote an o.n. set in H such that $H_n = \text{span}\{e_1,\ldots,e_n\}$. Similarly, $\{u_i, i \geq 1\}$ will denote an o.n. set such that $H_{Wn} = \text{span}\{u_1,\ldots,u_n\}$. Note that one of these sets is complete for H if and only if the other is complete; completeness is equivalent to $\|\tilde{P}_n x - x\| \to 0$ for all x in H.

Since V_n is an isometry from H_{W_n} to H_n, it follows that $V_n u_i = e_i$ for $i \leq n$. This is obviously true for $n = 1$; suppose that it holds for $n = K$. Then, since $H_{W_K} \subset H_{W_{K+1}}$ and $H_K \subset H_{K+1}$, the statement must hold for $n = K+1$, thus for all values of n.

Let n be fixed and define $G_n(\lambda) = \frac{1}{n}$ [# eigenvalues of $S_n < \lambda$]. G_n is a left-continuous non-negative step function, bounded above by 1. The family

$\{G_n, n \geq 1\}$ will be seen to completely define (for a given set of constraints) the coding capacity. The importance of $\sigma_{ess}(S)$ to characterizing $\{G_n, n \geq 1\}$ is demonstrated in Prop. 2 below. First, let $V[S,(H_n)]$ be the set of all γ in $\mathbb{R}$ such that for any K and any $\epsilon > 0$, there exists $n \geq K$ such that the number of elements in the sequence (β_i^n) satisfying $|\beta_i^n - \gamma| < \epsilon$ is $\geq K$.

To see that the set $V[S,(H_n)]$ determines $\overline{\lim}_n \frac{1}{n}$ [# eigenvalues of $S_n < \lambda$], let λ be fixed. Then $\overline{\lim}_n G_n(\lambda) = \overline{\lim}_n \frac{1}{n} \Sigma_{i=1}^K$ [# eigenvalues of S_n in $[\lambda_i,\lambda_{i+1})$], where $-1 = \lambda_1 < \ldots < \lambda_{K+1} = \lambda$. If $\overline{\lim}_n G_n(\lambda) > 0$, then there exists at least one $\lambda_i < \lambda$ such that $\overline{\lim}_n \frac{1}{n}$ [# eigenvalues in $[\lambda_i,\lambda_{i+1})$] > 0. This requires that $[\lambda_i,\lambda_{i+1})$ contain at least one point in $V[S, (H_n)]$. Thus, for any λ, $\overline{\lim}_n G_n(\lambda)$ is determined entirely by those $\gamma < \lambda$ such that $\overline{\lim}_n \frac{1}{n}$ [# eigenvalues of S_n in $(\gamma-\epsilon, \gamma+\epsilon)$] > 0 for every $\epsilon > 0$, and any such γ must belong to $V[S, (H_n)]$.

<u>Prop. 2 [6]</u>.

(1) Suppose that $\{u_n, n \geq 1\}$ is a c.o.n. set for H. Then $\sigma_{ess}(S) \subset V[S,(H_n)]$.

(2) Let θ_L and θ_U denote the smallest and largest points in $\sigma_{ess}(S)$. Then $V[S,(H_n)] \subset [\theta_L,\theta_U]$.

The results of Prop. 2 might lead one to hypothesize that $V[S,(H_n)] \subset \sigma_{ess}(S)$, that $V[S,(H_n)] = \sigma_{ess}(S)$ when $\{u_n, n \geq 1\}$ is complete, and that $\overline{\lim}_n \frac{1}{n}$ [# eigenvalues of $S_n < \lambda$] is independent of the choice of any c.o.n.s. $\{u_n, n \geq 1\}$. These three properties would be very useful. Unfortunately, all three are false, in general, although the first two will hold for some important choices of (H_n).

Prop. 3 [6].

(1) If $\{u_n, n \geq 1\}$ is not complete for H, then $V[S,(H_n)] \cap \sigma_{ess}(S) = \phi$ is possible.

(2) If $\{u_n, n \geq 1\}$ is complete for H, then:

(a) $V[S,(H_n)] \subset \sigma_{ess}(S)$ is not always true.

(b) Let $Q_\Delta(S,(H_n))$ be the eigenvalues β_i^n of S_n such that $|\beta_i^n - x| \geq \Delta > 0$ for all x in $\sigma_{ess}(S)$. Then $\overline{\lim} \frac{1}{n} Q_\Delta(S,(H_n))$ can be strictly positive.

(c) If $\{u_n, n \geq 1\}$ is complete for H, then $\overline{\lim}_n \frac{1}{n}$ {# eigenvalues of $S_n > \lambda\}$ need not be independent of the choice of $\{u_n, n \geq 1\}$.

Coding Capacity

The following theorems give a general result for coding capacity. See [6] for the proof of Theorem 1 and the Corollary.

Theorem 1.

(1)
$$\begin{aligned}
\overline{\lim_{n\to\infty}} \tfrac{1}{2} \int_{-1}^{B_n} \log\left[\frac{B_n+1}{\lambda+1}\right] dF_n(\lambda) &\leq \overline{\lim_{n\to\infty}} \frac{1}{n} C_W^n(nP) \\
&\leq \overline{\lim_{n\to\infty}} \tfrac{1}{2}\left[\int_{-1}^{B_n} \log\left[\frac{B_n+1}{\lambda+1}\right] dF_n(\lambda) + \frac{1}{n} H_{GN}^n(N)\right] \\
&\leq \overline{\lim_{n\to\infty}} \tfrac{1}{2} \int_{-1}^{B_n} \log\left[\frac{B_n+1}{\lambda+1}\right] dF_n(\lambda) + \overline{\lim_{n\to\infty}} \frac{1}{n} H_{GN}^n(N)
\end{aligned}$$

where B_n is defined by $\quad P = \int_{-1}^{B_n} (B_n - \lambda) dF_n(\lambda)$

and $\quad F_n(\lambda) = \frac{1}{n}$ {# eigenvalues of $S_n \leq \lambda\}$.

(2) If $\overline{\lim_n} \frac{1}{n} C_W^n(nP) > 0$, and $\lim_{n\to\infty} \frac{1}{n}$ {# eigenvalues of $S_n < \lambda$} exists for all λ in $\mathbb{R}$, then

$$\tfrac{1}{2} \int_{-1}^{B} \log\left[\frac{1+B}{1+\lambda}\right] dF(\lambda) \le \overline{\lim_n} \frac{1}{n} C_W^n(nP)$$

$$\le \tfrac{1}{2} \int_{-1}^{B} \log\left[\frac{1+B}{1+\lambda}\right] dF(\lambda) + \overline{\lim_n} \frac{1}{n} H_{GN}^n(N)$$

where F is a distribution function defined by

$$F(\lambda) = \lim_{n\to\infty} \frac{1}{n}\{\text{\# eigenvalues of } S_n \le \lambda\} = \lim_{n\to\infty} F_n(\lambda),$$

and the constant B is defined by

$$P = \int_{-1}^{B} [B-\lambda] dF(\lambda).$$

(3) If $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$, then $\overline{\lim_n} \frac{1}{n} C_W^n(nP) = 0$ if and only if $\overline{\lim_{n\to\infty}} \frac{1}{n}$ {# eigenvalues of $S_n < \lambda$} = 0 for all λ in $\mathbb{R}$. This requires that S be unbounded and occurs in particular if $+\infty$ is the only limit point of $\sigma(S)$.

Remarks. (1) In part (2), the same result holds if $F(\lambda) \equiv \lim_{n\to\infty} F_n(\lambda)$ exists for all $\lambda < B$, where B is defined by $P = \int_{-1}^{B}[B-\lambda]dF(\lambda)$.

(2) In the statement of Theorem 1, the probability distributions $\{F_n, n \ge 1\}$ could be replaced by $\{G_n, n \ge 1\}$, where $G_n = \frac{1}{n}$[# eigenvalues of S_n strictly less than λ]. This follows because the integrands are continuous functions, and are zero at the upper limit of the integral.

Corollary. If $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$, then bounds on $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$ are given by

$$\tfrac{1}{2} \log(1 + P/\lambda_{max}) \le \overline{\lim_n} \frac{1}{n} C_W^n(nP) \le \tfrac{1}{2} \log(1 + P/\lambda_{min})$$

where λ_{min} is the smallest limit point in the spectrum of S, and λ_{max} is the largest. Moreover, these bounds can be attained by proper choice of (H_n).

An alternative form of Theorem 1 can be given, as follows.

Theorem 1A. Suppose that B is the largest number such that $P \geq \overline{\lim_n} \int_{-1}^{B} [B_n-\lambda] dF_n(\lambda)$. Then

$$\overline{\lim_n} \int_{-1}^{B} \log\left[\frac{B_n+1}{\lambda+1}\right] dF_n(\lambda) \leq \overline{\lim_n} \frac{1}{n} C_W^n(nP)$$

$$\leq \overline{\lim_n} \left[\int_{-1}^{B} \log\left[\frac{B_n+1}{\lambda+1}\right] dF_n(\lambda) + \frac{1}{n} H_{GN}^n(N)\right]$$

where (F_n) and (B_n) are defined as in Theorem 1. If no such B exists, and $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$, then $\overline{\lim_n} \frac{1}{n} C_W^n(nP) = 0$.

The following result, together with Theorem 1 (and 1A), gives the coding capacity. Part (b) of this theorem can be proved from first principles, beginning with Feinstein's Lemma. However, Theorem 1 enables a proof to be given based upon results due to Ebert [10] and Gallager [11]. This approach not only shortens the exposition, but also provides error bounds.

Theorem 2. Let $C_W^\infty(P)$ be the coding capacity.

(a) $C_W^\infty(P) \leq \overline{\lim_n} \frac{1}{n} C_W^n(nP)$; when μ_N is Gaussian and $C_W^\infty(P) = 0$, then the maximum decoding error probability is equal to one for any rate > 0 (i.e., $\liminf_n \epsilon_n = 1$).

(b) If μ_N is Gaussian, then $C_W^\infty(P) = \overline{\lim_n} \frac{1}{n} C_W^n(nP)$.

(c) Suppose that μ_N is Gaussian. Then, for any fixed n, the maximum decoding error probability ϵ_n is bounded by

$$\epsilon_n \leq \left[\frac{2e^{s\delta}}{\gamma_n}\right]^2 \exp\left[-\tau[R(B[r,\rho])]\right]$$

where $0 \leq \rho \leq 1$, $s = \rho/(2[1+\rho]^2(1 + B[n,\rho])$, δ can be taken equal to $1/s$, and γ_n is an integration-normalizing constraint [11]. $B[n,\rho]$ is defined by

$$P = \frac{1}{n} \sum_{i=1}^{N(n,\rho)} \frac{(1+\rho)^2(B[n,\rho]-\beta_i^n)(B[n,\rho]+1)}{(1+\rho)(1+B[n,\rho]) - \rho(1+\beta_i^n)}$$

$$N(n,\rho) = \sup\{i \leq n: \beta_i^n < B[n,\rho]\}$$

and

$$\tau[R(B[n,\rho])] = \frac{\rho n P}{2(1+B[n,\rho])(1+\rho)} + \tfrac{1}{2} \sum_{i=1}^{N(n,\rho)} \log\left[\frac{1+B[n,\rho]}{(1+\rho)(1+B[n,\rho]-\rho(1+\beta_i^n)}\right].$$

The corresponding number of code words is $[e^{R(B[n,\rho])}]$, where

$$R(B[n,\rho]) = \tfrac{1}{2} \sum_{i=1}^{N(n,\rho)} \log\left[\frac{B[n,\rho] + 1}{\beta_i^n + 1}\right].$$

<u>Proof</u>. The fact that $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$ is always an upper bound on coding capacity follows by a standard application of Fano's inequality; see, for example, p. 168 of [16]. The resulting inequality for a code (k_n, n, ϵ_n) is $\epsilon_n \geq 1 - \frac{C_W^n(nP) + \log 2}{\log k_n}$. This gives $\overline{\lim_n} \frac{1}{n} C_W^n(nP)$ as an upper bound on capacity. Suppose now that μ_N is Gaussian, and assume that (b) of the theorem holds. If $C_W^\infty(P) = 0$, then for any positive number R, $\frac{1}{n} C_W^n(nP) < R$ for all sufficiently large n, so that $\overline{\lim_n} \frac{1}{n} C_W^n(nP) = 0$, giving $\epsilon_n \to 1$ for any positive rate R.

To obtain the lower bound for $C_W^{\infty}(P)$ when μ_N is Gaussian, one can apply the results of Ebert [10], [11]. They involve a vector channel, consisting of a set of n parallel one-dimensional Gaussian channels where the noises in the different channels are mutually independent with variances β_i^n, $i \le n$. The code words are vectors x which satisfy the constraint $\Sigma_1^n x_i^2 \le nP$. With this model, Ebert shows the existence of a code $([e^{R(B[n,\rho])}], n, \epsilon_n)$, with $R(B[n,\rho])$ defined as in (c) of the theorem, and ϵ_n obeying the upper bound given there. In those equations, $\rho = 0$ gives $\tau[R(B(n))] = 0$ and $P = \frac{1}{n}\sum_{i=1}^{N(n)} [B(n) - \beta_i^n]$. $\tau[R(B[n,\rho])] > 0$ for $\rho > 0$, and $B[n,\rho]$ decreases (for fixed P) as ρ increases.

Thus, for every $\rho > 0$, one has an admissible rate R_ρ, defined by

$$R_\rho = \overline{\lim_n} \frac{1}{n} R(B[n,\rho]) = (\text{by Theorem 1})\ \overline{\lim_n}\ \tfrac{1}{2} \int_{-1}^{B[n,\rho]} \log\left[\frac{B[n,\rho]+1}{\lambda + 1}\right] dF_n(\lambda).$$

This gives a lower bound on capacity of

$$\overline{\lim_{\rho\to 0}} R_\rho = \overline{\lim_{\rho\to 0}}\ \overline{\lim_n}\ \tfrac{1}{2} \int_{-1}^{B[n,\rho]} \log\left[\frac{B[n,\rho]+1}{\lambda + 1}\right] dF_n(\lambda).$$

Since $\int_0^{B[n,\rho]} \log\left[\frac{B[n,\rho]+1}{\lambda + 1}\right] dF_n(\lambda)$ is non-decreasing as ρ decreases for fixed n, it follows that

$$\overline{\lim_{\rho\to 0}} R_\rho \le \overline{\lim_n}\ \tfrac{1}{2} \int_{-1}^{B[n,0]} \log\left[\frac{B[n,0]+1}{\lambda + 1}\right] dF_n(\lambda).$$

Moreover, if for $\Delta_n > 0$, $B[n,0] + 1 - \Delta_n > 0$, $B[n,\rho] = B(n,0) - \Delta_n$, then

$$\overline{\lim_n} \int_{-1}^{B[n,0]} \log\left[\frac{B[n,0]+1}{\lambda + 1}\right] dF_n(\lambda) - \overline{\lim_n} \int_{-1}^{B[n,\rho]} \log\left[\frac{B[n,\rho]+1}{\lambda + 1}\right] dF_n(\lambda)$$

$$\le \overline{\lim_n} \int_{-1}^{B[n,0]} \log\left[\frac{B[n,0]+1}{B[n,\rho]+1}\right] dF_n(\lambda) \le \overline{\lim_n} \frac{\Delta_n}{B[n,0] + \Delta_n - 1}.$$

and since B[n,0] is bounded away from -1, we obtain

$$\overline{\lim_{\rho\to 0}}\, R_\rho = \overline{\lim_{n}}\ \tfrac{1}{2} \int_{-1}^{B[n,0]} \log\left[\frac{B[n,0]+1}{\lambda + 1}\right] dF_n(\lambda)$$

where $\quad P = \frac{1}{n} \sum_{i=1}^{N(n)} [B[n,0] - \beta_i^n] \quad$ and $\quad N(n) = \sup\{i \leq n: \beta_i^n \leq B[n,0]\}$.

To apply this result, one proceeds similarly to [11]. For fixed n, the channel considered here has code words in H_n constrained by $\|x\|^2_{W,n} = \|y\|^2 \leq nP$, where y is the unique element of H_n satisfying $R^{1/2}_{W,n}y = x$. The noise μ^n_N has covariance operator $R_{N,n} = R^{1/2}_{W,n}(I_n + S_n)R^{1/2}_{W,n}$. Thus, this is the same as the H_n channel with code words $y_1,\dots,y_k$ satisfying $\|y\|^2 \leq nP$ and with the additive noise having covariance $I_n + S_n$. Expanding all code words and noise sample paths in terms of the orthonormal eigenvectors of $I_n + S_n$, one obtains a channel whose output is the sum of a vector of n independent parallel Gaussian channels, with the outputs of the n channels being mutually-orthogonal elements of H_n. The i^{th} channel has additive noise with covariance operator $(1 + \beta_i^n)v_i^n \otimes v_i^n$, and the code words (y_k) can be written as $y_k = \Sigma^n_{i=1} y_{ki}v_i^n$, where $\Sigma^n_{i=1}\ y^2_{ki} \leq nP$, $y_{ki}v_i^n$ being the input to the i^{th} parallel channel when the code word y_k is selected. Since the individual channels have outputs that are mutually orthogonal in H_n, the probability of correct decoding for the summed output is the probability that all of the individual channel outputs are correctly decoded. The Ebert results thus apply, and part (b) of the lemma is proved. Part (c) follows. □

Applications: Discrete-Time Channels

For the discrete-time memoryless Gaussian channel with $R_W = I$, the theorems give easily-obtained new results for nonstationary channels. In this case, $R_N = I+S$; since S is diagonal, the eigenvalues of $I_n + S_n$ are (α_i), $i \leq n$, where $I+S = \mathrm{diag}[\alpha_1,\alpha_2,\dots]$. The spectral limit points $\{\theta_1,\dots,\theta_K\}$ of

$I+S$ are the limit points of the eigenvalues (α_i) of R_N. These limit points and their "relative frequencies" completely characterize the capacity for this simple channel, whereas in the general case the family of distribution functions (F_n) can converge to a distribution function with points of increase at points that are not limit points of the spectrum of $I + S$. For the stationary memoryless discrete-time channel with $R_W = I$ and $R_N = \sigma^2 I$, σ^2 is the only limit point of $I + S$, and so by Prop. 2 one obtains the well-known result that $C_W^\infty(P) = \frac{1}{2}\log\left[1 + \frac{P}{\sigma^2}\right]$. This is also the value of $C_W^\infty(P)$ if $R_W = I$ and $R_N = \sigma^2 I + M$, where M is any operator in H such that M is compact. This follows from the fact that compact operators in a Hilbert space are exactly those operators that have zero as the only limit point of their spectrum. Thus, if the noise is of the form $N = N_1 + N_2$, N_1 stationary and uncorrelated with variance σ^2, and N_2 independent of N_1 with $\underset{\sim}{E}_2\left[\sum_{n\geq 1} x_n^2\right] < \infty$, then the coding capacity is again $\frac{1}{2}\log\left[1 + \frac{P}{\sigma^2}\right]$. Of course, we are assuming as always that all processes have zero mean.

These remarks follow from the following result.

<u>Prop. 4</u>. Let $H = \ell_2$, $H_n = \{x\colon x_i = 0 \text{ for } i \geq n\}$, $R_W = I$, and $R_N = \operatorname{diag}[\sigma_i, i \geq 1]$. Suppose that μ_N is Gaussian and that (σ_i) has the limit points $\theta_1 < \theta_2 < \dots < \theta_K$. Then

$$C_W^\infty(P) = \overline{\lim_n} \sum_{j=1}^{J} \gamma_j^n \log\left[\frac{B_n}{\theta_j}\right]$$

where $\gamma_j^n = M_j^n/n$, M_j^n is the number of elements in the sequence $(\sigma_i, i \leq n)$ belonging to $(\theta_j-\epsilon, \theta_j+\epsilon)$, $\epsilon > 0$ satisfies $2\epsilon < \min\{\theta_{j+1}-\theta_j\colon j \geq 0, \theta_0 \equiv 0\}$, (B_n) is defined by $P = \Sigma_{j=1}^{J}\gamma_j^n(B_n - \theta_j)$, and J is the largest integer $\leq K$ such that

$$P \geq \overline{\lim_{n}} \sum_{i \leq J} \gamma_i^n(\theta_J - \theta_i).$$

If $\lim_{n\to\infty} \gamma_i^n \equiv \gamma_i$ exists for $i \leq J$ with J as defined below, then

$$C_W^\infty(P) = ½ \sum_{i=1}^{J} \gamma_i \log\left[\frac{B}{\theta_i}\right]$$

where J and B are defined by $P = \Sigma_{j=1}^{J}\gamma_j(B - \theta_j)$, with θ_J the largest element of $\{\theta_1,\ldots,\theta_K\}$ satisfying $P \geq \Sigma_{i=1}^{K}\gamma_i(\theta_J - \theta_i)$.

Proof. Direct application of the theorems.

The case where (σ_n) has an infinite set of limit points is presumably of marginal interest; the capacity in that case is less easy to visualize, but is also obtained immediately from the theorems.

Heuristically, one can view this channel as equivalent to K parallel independent discrete-time memoryless channels. The kth channel has non-zero noise components only for those indices j such that $|\sigma_j - \theta_k| < \epsilon$. For fixed n, a code word for the composite channel is then given by $y = (y_1, y_2, \ldots, y_K)$, where y_k is the component of the code word for the kth channel and must satisfy $y_{kj} = 0$ if $|\sigma_j - \theta_k| \geq \epsilon$, $y_{k_j} = 0$ if $j > n$, while $\Sigma_{k=1}^{K}\Sigma_{j\geq 1} y_{kj}^2 \leq nP$.

The effect of Prop. 4 is then to replace the original channel by K parallel independent channels, the kth channel being a memoryless stationary channel with noise variance θ_k. The coder then uses the K channels according to the probability distributions (γ_i^n), the (θ_i), and P.

Using this viewpoint, the results of Prop. 4 then show that the coder chooses his code word as $y = (y_1, y_2, \ldots, y_K)$, where y_k is the component of the code word that is used as an input to the kth channel. For fixed n, he chooses the code word y according to the constraint $\frac{1}{n} \Sigma_{j=1}^{n} y_{k_j}^2 \leq \gamma_k^n(B_n - \theta_k)$.

$k \le J$, with $y_{k_j} = 0$ if $j > n$ or $|\sigma_j - \theta_k| \ge \epsilon$, and $\Sigma_{j=1}^{n} y_{k_j}^2 = 0$, $k > J$. In the case where $\lim_{n\to\infty} \gamma_k^n \equiv \gamma_k$ exists for $k \le J$, this gives the constraint $\frac{1}{n} \Sigma_{j=1}^{n} y_{k_j}^2 \le \gamma_k(B - \theta_k)$, $k \le J$, $\Sigma_{j=1}^{n} y_{k_j}^2 = 0$, $k > J$, where $B = [P + \Sigma_{k=1}^{J} \gamma_k \theta_k]/\Sigma_{j=1}^{J} \gamma_j$. The capacity is then $C_W^{\infty}(P) = \Sigma_{j=1}^{J} \gamma_j C_j^{\infty}(P)$, where $C_j^{\infty}(P)$ is the capacity of the jth channel subject to the above constraint.

If the sequence of noise variances $(\sigma_i,\ i \ge 1)$ consists only of numbers in the finite set $\{\theta_i,\ i \le K\}$, then Prop. 4 shows that even for a noise variance θ_i that is repeated infinitely often, such a component of the noise will have no effect on capacity if the relative frequencies (γ_i^n) are such that $\overline{\lim_n}\, \gamma_i^n = 0$. However, if P is small and a limit point θ_i of $(\sigma_k,\ k \ge 1)$ is so large that it does not appear in the expression for capacity (i.e., $i > J$ as given in Prop. 4), then this limit point may still affect the capacity if $\overline{\lim_n}\, \gamma_i^n > 0$, and will always affect capacity if $\liminf_n \gamma_i^n > 0$. This may be viewed as somewhat unexpected, since such a θ_i would represent one of the "K parallel channels" for which the effective input is zero. However, this is a point where the heuristic "parallel channel" analogy breaks down; this is due to the fact that the "i^{th} channel" is present for a fraction γ_i^n of the available time, n, and the coder is defining capacity in terms of transmission time (i.e., part of the allowable dimensionality is being used by a "channel" which conveys no information).

Prop. 4 (and the theorems, for more general channels) has obvious applications to some multiuser channels. For example, consider a time-division multiaccess Gaussian channel defined as follows. There are K sources. For transmission up to time $t = n$, with $n \ge 1$, the j^{th} source uses the channel a fraction $\gamma_j^n = n(j)/n$ of the time. The noise added to the code word of the j^{th} source has variance θ_j^n (for the $n(j)^{th}$ transmission by the j^{th}

source) when the overall transmission time is n. The sources have the overall constraint $\frac{1}{n}\Sigma_{j=1}^{K}\Sigma_{i=1}^{n(j)}y_{ji}^{2} \leq P$ for each $n \geq 1$. Prop. 4 then shows how the available power should be allocated among the K sources, and gives the capacity. Examples of such channels include that defined by an earth-orbiting satellite with K widely-separated ground-based transmitters, or a channel where K sources feed into a central relay station.

This example is for a very simple case of multiaccess channels. More general problems can be analyzed. However, the basic idea is the same; one identifies a source (or group of sources) with a limit point (or set of limit points) in $\sigma(S)$, and the corresponding (γ_i^n), $n \geq 1$, is the fraction of time the source uses the channel up to time n. This is for the memoryless discrete-time channel; the theorems can be used to analyze more general channels. A particular aspect of this model is that the fraction of time that each source uses the channel can vary with time; similarly for the noise environment faced by each source.

One can also use Prop. 4 (and the theorems) to analyze jamming channels. For example, if a jammer must vary his energy over different time periods, Prop. 4 will permit the calculation of capacity for a given set of $\{\underline{\gamma}, \underline{\theta}, P\}$. More generally, as will be discussed elsewhere, the theorems permit one to determine the jammer's minimax strategy, subject to various types of constraints on the jamming signal.

It can be seen that the best choice of (H_n) from the viewpoint of maximizing capacity will be the natural choice $H_n = \{x: x_i = 0, i > n\}$ only in special situations. If $\theta_1 < \theta_2 < \ldots < \theta_K$ are the limit points of the noise variances, and μ_N is Gaussian, then an optimum choice of (H_n) is given by $H_n = A_n \cap B$, where

$$A_n = \{x: x_i = 0, i > k(n)\},$$

$$B = \{x: x_i = 0, |\sigma_i - \theta_1| \geq \epsilon\},$$

$k(n)$ = smallest integer such that $|\sigma_i - \theta_1| < \epsilon$ for exactly n values of $i \leq k(n)$,

and

$$\epsilon < \theta_2 - \theta_1.$$

This definition of (H_n) gives

$$C_W^\infty(P) = \tfrac{1}{2}\log[1 + P/\theta_1],$$

which is the maximum possible value for a given S. Of course, this squares perfectly with intuition; the original channel is transformed into a "channel" having limiting noise variance θ_1, at the expense of increasing the transmission time required to achieve a specified decoding error.

The choice of $H_n = \{x: x_i = 0, i > n\}$ gives the optimum (H_n) when μ_N is Gaussian if and only if $\limsup_n \gamma_1^n = 1$. Thus, if choice of (H_n) is part of the system design, then it is only in this case that the capacity is equal to that which is obtained in the classical Gaussian channel. Conversely, the classical channel gives the worst possible choice of (H_n) if and only if $\overline{\lim_n}\, \gamma_i^n = 0$ for all $i \leq K-1$. For example, consider a channel with noise variances (α_i) given as follows:

$$\begin{aligned} \alpha_i &= 2 && i = j^2,\ j \text{ any integer} \\ &= 2000 && \text{otherwise}. \end{aligned}$$

Then $\theta_1 = 2$, $\theta_2 = 2000$; if (H_n) is defined by $H_n = \{x: x_i = 0, i > n\}$, then $\overline{\lim_n}\, \gamma_1^n = 0$, so that $C_W^\infty(P) = 0$. However, consider (H_n) defined by

$$H_n = \{x: x_i = 0, i > n^2\} \cap B, \text{ where}$$

$$B = \{x: x_i = 0, i \neq j^2 \text{ for any integer } j\}.$$

If μ_N is Gaussian, then this is an optimum choice of (H_n) and gives $C_W^\infty(P) = \frac{1}{2} \log[1 + P/2]$. One should notice the differences in definition of capacity. However, examples that are even more striking can be constructed for channels with memory; it is possible to use a code word set of arbitrarily large cardinality with maximum decoding error going to zero as transmission time $n \to \infty$, even though the classical channel has zero capacity. This can be seen from Theorem 1.

In the above analyses, the cost of transmission time is not quantified. In the classical channel, one is implicitly assuming that transmission time must be minimized for a code word set of dimension n. The formulation given here permits one to remove this constraint. When μ_N is Gaussian, Theorem 2 can be used to determine tradeoffs between the transmission time, maximum decoding error, and cardinality of the code word set, for each n, for various choices of (H_n).

As a final remark on the memoryless channel, one may note that for the stationary nonGaussian channel with noise variance σ^2, Shannon obtained the result [18] that for $R_W = I$,

$$\lim_{n\to\infty} \frac{1}{n} C_W^n(nP) \leq C(P,G) + \frac{1}{2} \log e^{2H(G)-2H(N)}$$

where G is the zero-mean Gaussian random variable with variance σ^2, C(P,G) is the capacity when G is the noise, and H denotes differential entropy. As previously seen,

$$\frac{1}{2} \log e^{2H(G)-2H(N)} = H_G(N) = \frac{1}{n} H_{GN}^n(N);$$

here $H_G(N)$ denotes the relative entropy of the random variable N to the r.v.

G; the last equality follows by the channel being memoryless.

The theorems also permit an immediate calculation of the known result for capacity of the stationary discrete-time channel with memory when a simple energy constraint is used.

Prop. 5. Let $H = \ell_2$, $R_W = I$, and $H_n = \{x \text{ in } \ell_2: x_i = 0, i > n\}$, with R_N given by a spectral density Φ. Then, if $\overline{\lim_n} \frac{1}{n} H^n_{GN}(N) = 0$,

$$\overline{\lim_n} \frac{1}{n} C^n_W(nP) = \tfrac{1}{2} \int_{\{x:\ \Phi(x) \leq B_0\}} \left[\log \frac{B_0}{\Phi(x)}\right] dx$$

where B_0 satisfies

$$P = \frac{1}{2\pi} \int_{\{x:\ \Phi(x) \leq B_0\}} [B_0 - \Phi(x)]dx.$$

For this application, the distribution function F is defined by $2\pi F(x) = m\{y: \Phi(y) \leq x\}$, where m is Lebesgue measure on $[-\pi,\pi]$.

Proof. In this case, $\tilde{P}_n = V_n$, so

$$G_n(\lambda) = \frac{1}{n}\{\text{number of eigenvalues of } \tilde{P}_n(I+S)\tilde{P}_n < \lambda + 1\}.$$

Now, by the Toeplitz distribution theorem [12]

$$\lim_{n\to\infty} G_n(\lambda) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{[0,\lambda)}(\Phi(x))dx = \frac{1}{2\pi} m\{x: \Phi(x) < \lambda\},$$

when m is Lebesgue measure, and the result follows from Theorem 1. □

Bounds on capacity of this channel can also be given. Suppose that $m \leq \Phi(x) \leq M$, $|x| \leq \pi$. If $\overline{\lim_n} \frac{1}{n} H^n_{GN}(N) = 0$, then $\tfrac{1}{2} \log\left[1 + \frac{P}{M}\right] \leq \overline{\lim_n} \frac{1}{n} C^n_W(nP) \leq \tfrac{1}{2} \log\left[1 + \frac{P}{m}\right]$. For μ_N Gaussian, the fact that $\overline{\lim_n} \frac{1}{n} C^n_W(nP) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{[0,B]}(\Phi(\lambda)) \log[\frac{B}{\Phi(\lambda)}]d\lambda$ with $P = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{[0,B]}(\Phi(\lambda))[B - \Phi(\lambda)]d\lambda$ has been known/assumed for many years. It is credited to Pinsker under the assumption of a stationary Gaussian signal [15]. However, it is apparently only recently that a complete proof has been given allowing a general nonstationary signal process [15]. In [15], it is assumed that Φ is continuous; that assumption is not needed here.

Applications: Continuous-Time Channels of Fixed Duration

In this section the code words and noise paths are elements of $L_2[0,T]$, where $T < \infty$ is fixed. The available energy per transmitted code word is P_0: for each fixed n and a given (H_n) and R_W, one has $\|x\|^2_{W,n} \leq P_0$ for each code word x. The question is whether or not arbitrarily small maximum decoding e-ror probability can be achieved by making P_0 arbitrarily large without limiting the cardinality of the code word set. It can be assumed that limiting the cardinality of the code word set is equivalent to limiting its dimensionality.

This problem is fundamentally different from that of the classical continuous-time channel, wherein the code words are limited to an energy of TP and T is permitted to become arbitrarily large. In the present case, for a fixed value of P_0, one can set $P_0 = nP$. Theorem 2 can then be used to determine an upper bound on the maximum decoding error probability. Of course, this requires that the eigenvalues (β^n_i) be determined for sufficiently many values of n, so that the expressions given in Theorem 2 can be evaluated. For a given value of $P_0 = nP$, one then determines $B[n,\rho]$ for suitable values of ρ (ρ in $(0,1)$) and chooses the values of n and ρ that give the most satisfactory compromise between the size of the code word set and the maximum decoding error probability.

In the balance of this section, attention will be focused on determining capacity. As previously discussed, a "rate" is the rate of increase in log[cardinality of the code word set] as a function of increasing dimensionality. R is then an admissible rate if the maximum decoding error can be made arbitrarily small by indefinitely increasing log[cardinality of the code word set] at the rate R. From Theorem 2, if μ_N is Gaussian and the capacity is zero, then the maximum probability of decoding error converges to one as the cardinality of the code word set becomes arbitrarily large, for any positive rate. We begin with an example illustrating this situation.

Let $H = L_2[0,T]$ and suppose that (N_t) has covariance operator having an inverse which is a densely-defined differential operator of order 2p. For example, if p = 1, N could have covariance function $e^{-\alpha(t-s)}$ $(\alpha > 0)$ or min(t,s). Let R_W be an integral operator whose inverse is a densely-defined differential operator of order $\geq 4p$. Thus, if p = 1, R_W could be defined by a covariance function corresponding to a spectral density which behaves, for $|\lambda| \to \infty$, as $\Phi_W(\lambda) \cong 1/\lambda^{4k}$, where $k \geq 1$. Then, if (N_t) is Gaussian, or more generally when $\overline{\lim_n} \frac{1}{n} H^n_{GN}(N) = 0$, the capacity is zero, regardless of the definition of the subspaces (H_n). This result follows from the fact that $R_N = R_W^{1/2}(I+S)R_N^{1/2}$, where I+S is the inverse of a compact operator, thus has a single limit point in its spectrum, equal to $+\infty$.

Some stationary channels with rational spectral densities defining both R_N and R_W constitute a special case of the above example. A complete set of results can be given for all stationary channels where R_N and R_W are defined by rational spectral densities.

<u>Prop. 6</u>. Suppose that $H = L_2[0,T]$. Let (N_t) be stationary and Gaussian with rational spectral density Φ_N, and suppose that R_W is defined by a rational spectral density Φ_W. Then, for any choice of (H_n):

a) the coding capacity will be non-zero if and only if

$$\lim_{|\lambda|\to\infty} \frac{\Phi_W(\lambda)}{\Phi_N(\lambda)} = \alpha > 0;$$

b) the capacity is given by $C_W^\infty(P) = \frac{1}{2} \log[1 + \alpha P]$.

This also gives an upper bound on the coding capacity if N is nonGaussian and the following conditions are satisfied: $(N_t) = (G_t + V_t)$ where (G_t) is stationary and Gaussian with rational spectral density Φ_G, (V_t) is independent of (G_t), possibly nonGaussian, stationary or nonstationary, and such that $\underset{\sim}{E} \int_{-\infty}^{\infty} \frac{|\hat{v}(\lambda)|^2}{\Phi_G(\lambda)} d\lambda < \infty$ for the sample paths v of (V_t), where $\hat{v}$ is the Fourier transform of v. Then, $C_W^\infty(P) = \frac{1}{2} \log\left[1 + P \lim_{|\lambda|\to\infty} \frac{\Phi_W(\lambda)}{\Phi_G(\lambda)}\right]$.

Proof. When R_W and R_N are defined by rational spectral densities Φ_W and Φ_N, then it can be shown from well-known results [2], [13] that $R_W = R_N^{1/2}(I+V)R_N^{1/2}$, where the operator V has the following properties:

a) V is bounded if and only if $\{\Phi_W(\lambda)/\Phi_N(\lambda),\ |\lambda| > 0\}$ is bounded;

b) if Φ_W/Φ_N is integrable over $(-\infty,\infty)$, then I+V is trace-class;

c) if V is bounded and

$$\lim_{|\lambda|\to\infty} \frac{\Phi_W(\lambda)}{\Phi_N(\lambda)} = \alpha,$$

then I+V has a single limit point for its spectrum, equal to α.

Using these facts, one notes that since $R_N = R_W^{1/2}(I+V)^{-1}R_W^{1/2} = R_W^{1/2}(I+S)R_W^{1/2}$, S must be unbounded with the single limit point $+\infty$ if Φ_W/Φ_N is integrable (since I+V has only zero as a limit point). By Theorem 1 and Prop. 2, $C_W^\infty(P)$ is then zero. If I+V has α as its only limit point, then $I + S = (I+V)^{-1}$ must have α^{-1} as its only limit point (note that I+V) + $(1-\alpha)I = V - \alpha I$ is compact); (b) of Prop. 6 follows. The remainder of Prop. 6 can be obtained from the results of [13]. □

As can be seen, there is no "waterfilling" aspect to the statement of Prop. 6. This is because the operator S has only a single limit point in its spectrum if $H = L_2[0,T]$ and R_W and R_N are defined by rational spectral densities. The waterfilling interpretation can be applied to Theorem 1; it is in terms of the family of pure jump distribution functions (F_n), which is determined from $V[S,(H_n)]$. It is notable that the results of Prop. 6 are independent of the value of T.

From (a) of Prop. 6, $C_W^\infty(P) = 0$ if Φ_W/Φ_N is integrable. This is the class of channels considered in the Holsinger-Gallager result for the classical continuous-time channel [11, Sec. 8.5], $T \to \infty$. It may be judged only natural that $C_W^\infty(P) = 0$ if T is fixed, since $\overline{\lim_{T\to\infty}} \frac{1}{T} C_W^T(TP)$ is finite, where $C_W^T(TP)$ is the capacity for the channel restricted to the interval $[0,T]$ and with the constraint $\underset{\sim}{E}\|X\|_{W,T}^2 \leq PT$. Since the dimensionality is not constrained in computing $C_W^T(TP)$, $C_W^T(TP) \geq C_W^n(nP)$ when $T = n$. Moreover, when T is fixed, the capacity is determined in terms of $(\frac{1}{n} C_W^n(nP))$. However, if R_W and R_N are given by rational spectral densities such that $\lim_{|\lambda|\to\infty} \frac{\Phi_W(\lambda)}{\Phi_N(\lambda)}$ is finite and non-zero, then the capacity is finite and non-zero for both the classical channel [7] and (by Prop. 6) the fixed-time channel.

Another result along these lines is the following.

<u>Prop. 7</u>. Suppose that $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$, that $\{u_n, n \geq 1\}$ is complete, and that μ_W is the zero-mean Gaussian probability with R_W as covariance. Then, $\overline{\lim_n} \frac{1}{n} C_W^n(nP) = ½ \log[1+P]$ if μ_W is absolutely continuous w.r.t.

(with respect to) μ_N. More generally, $\overline{\lim_n} \frac{1}{n} C_W^n(nP) = \frac{1}{2} \log[1 + \frac{P}{1-\alpha}]$ if there exists a constant $\alpha < 1$ such that $R_N^\alpha \equiv R_N + \alpha R_W$ is strictly positive definite and μ_W and $\mu_N^{G,\alpha}$ are mutually absolutely continuous, where $\mu_N^{G,\alpha}$ is the zero-mean Gaussian probability with covariance operator R_N^α.

<u>Proof</u>. If $\mu_W \sim \mu_N^{G,\alpha}$, then $R_N + \alpha R_W = R_W^{1/2}(I+T)R_W^{1/2}$, T Hilbert-Schmidt, so that $R_N = R_W^{1/2}(I + [T-\alpha I])R_W^{1/2}$; since $T - \alpha I \equiv S$ has $-\alpha$ as the only limit point of $\sigma(S)$, the result follows. Note that $\alpha < 1$ is necessary because $I + T - \alpha I$ must be non-negative, requiring $T \geq (\alpha-1)I$ Since T is compact, this can hold only for $\alpha \leq 1$, and the case $\alpha = 1$ violates the basic (and necessary, for finite capacity) assumption that $R_N = R_W^{1/2}(I+S)R_W^{-1/2}$ with $(I+S)^{-1}$ bounded. For, if $\alpha = 1$, then $R_N + \alpha R_W = R_W + R_W^{1/2}TR_W^{1/2}$ and T Hilbert-Schmidt implies that $R_N = R_W^{1/2}TR_W^{-1/2}$, and T^{-1} cannot be bounded. □

The coding capacity of the matched channel ($R_W = R_N$) is $\frac{1}{2}\log(1+P)$. Thus, if $\mu_W \sim \mu_{GN}$, then the difference between the two operators R_W and R_N is not sufficient to affect the coding capacity. However, if $\mu_W \perp \mu_{GN}$, then one still obtains finite capacity under the assumptions of Prop. 7, but its value can be greater or smaller than that of the matched channel.

<u>Prop. 8</u>. Suppose that $\overline{\lim_n} \frac{1}{n} H_{GN}^n(N) = 0$. In order that $C_W^\infty(P)$ be more than zero, it is necessary that $R_W = R_N^{1/2}TR_N^{1/2}$ with T bounded but *not* compact.

Prop. 8 follows immediately from the preceding. $(I+S)^{-1} = T$, so that T compact implies that (I+S) has $+\infty$ as the only limit point of its spectrum. This is actually the situation that holds in the Holsinger-Gallager model.

The key to interpreting these results lies in the expressions for information capacity when μ_N is Gaussian [5]. When S has a single limit point equal to $+\infty$ (as in the case of Example 2 and in Prop. 6 when Φ_W/Φ_N is integrable), then S has a CONS of eigenvectors and corresponding eigenvalues (λ_i), $\lambda_i \nearrow \infty$. With no dimensionality constraint on the transmitted signal process, Theorem 3 and Corollary 4 of [5] show that the optimum signal process (for achieving information capacity) has finite-dimensional support when $P_0 \equiv nP$ is fixed.

It then follows that increasing the dimensionality of the signal space past the optimum value actually decreases information capacity. This is consistent with $\overline{\lim}_{n\to\infty} \frac{1}{n} C_W^n(nP) = 0$ for any fixed value of P.

However, when the smallest limit point θ of S is finite, then for a sufficiently large value of $P_0 \equiv nP$, one will have $P_0 + \sum_{i=1}^{K} \lambda_i \geq K\theta$, where now (λ_i) denotes those (increasing) eigenvalues of S strictly less than θ. Theorem 2c of [5] then applies, setting the dimensionality as n = M. Let K = min(L,M), where L is the number of eigenvalues of S strictly less than θ. Then, permitting H_M to be any M-dimensional subspace, one has [5]

$$C_W^M(P_0) = \tfrac{1}{2} \sum_{i=1}^{K} \log\left[\frac{1+\theta}{1+\lambda_i}\right] + \frac{M}{2} \log\left[1 + \frac{P_0 + \sum_{i=1}^{K} (\lambda_i - \theta)}{M(1+\theta)}\right]$$

with the constraint given by $E_{\mu_X} \|x\|_W^2 \leq P_0$. Since $P_0 = MP$, taking $M \to \infty$ gives $\lim_{M\to\infty} \frac{1}{M} C_W^M(MP) = \tfrac{1}{2} \log\left[1 + \frac{P}{1+\theta}\right]$.

Of course, this is an upper bound for the coding capacity, in general, since (H_M) can be any sequence such that H_M is M-dimensional, including sequences that are not ordered by inclusion. However, if S has only one limit point θ in its spectrum (as in Prop. 6 when $\Phi_W(\lambda)/\Phi_N(\lambda) \to \alpha \neq 0$) then

$\frac{1}{2} \log\left[1 + \frac{P}{1+\theta}\right]$ is in fact the coding capacity, as has been seen from the corollary to Theorem 1.

The preceding results thus enable one to determine whether or not arbitrarily small error probability can be achieved while indefinitely increasing the log of the cardinality of the code word set at some positive rate, and give the capacity. As discussed, the capacity in this framework is the supremum of all admissible rates, and the "rate" is defined as $[\log k_n]/n$, where n is the dimensionality of the code word set and k_n is the number of code words.

[17] F. Riesz and B. Sz-Nagy, *Functional Analysis*, Ungar, New York, 1955.

[18] C.E. Shannon, A Mathematical Theory of Communication, *Bell System Tech. J.*, 27, 623-656 (1948).